Working Papers

Federal Reserve Bank of Boston

No. 07-10

Space and Time in Macroeconomic Panel Data: Young Workers and State-Level Unemployment Revisited

Christopher L. Foote

Abstract:

A provocative paper by Shimer (2001) finds that state-level youth shares and unemployment rates are negatively correlated, in contrast to conventional assumptions about demographic effects on labor markets. This paper updates Shimer's regressions and shows that this surprising correlation essentially disappears when the end of the sample period is extended from 1996 to 2005. This shift does not occur because of a change in the underlying economy during the past decade. Rather, the presence of a cross-sectional (that is, spatial) correlation in the state-level data sharply reduces the precision of the earlier estimates, so that the true standard errors are several times larger than those originally reported. Using a longer sample period and some controls for spatial correlation in the regression, point estimates for the youth-share effect on unemployment are positive and close to what a conventional model would imply. Unfortunately, the standard errors remain very large. The difficulty of obtaining precise estimates with these data illustrates a potential pitfall in the use of regional panel data for macroeconomic analysis.

JEL Classifications: J64, E24, J11

Christopher L. Foote is a Senior Economist and Policy Advisor at the Federal Reserve Bank of Boston. His email address is chris.foote@bos.frb.org.

This paper, which may be revised, is available on the web site of the Federal Reserve Bank of Boston at http://www.bos.frb.org/economic/wp/index.htm.

Thanks are due to Robert Shimer, John Driscoll, Gábor Kézdi, Gary Solon, James Stock, Yolanda Kodrzycki, Paul Willen, and seminar participants at the Boston Fed, William and Mary, Duke, Harvard, and the 2007 Federal Reserve System Regional conference for useful comments. Shimer also kindly supplied much of the birth-rate data, and James LeSage made useful MATLAB code available on his website. Christopher Goetz provided outstanding research assistance.

This version: September 26, 2007

1. Introduction

Macroeconomists often use state-level or regional data when national variation is insufficient or when a particular identification strategy is feasible only on a sub-national level. A recent example is a careful and provocative paper by Robert Shimer (2001), who investigates the effects of demographic change on labor markets. The traditional demographic adjustment for the unemployment rate assumes that aggregate unemployment moves mechanically along with the population shares of various demographic groups.[1] For example, the increase in young workers in the 1970s and 1980s is generally thought to explain part of the increase in overall U.S. unemployment during that time, because young workers experience higher unemployment rates than older workers.[2] In his paper, Shimer uses data from U.S. states to estimate — rather than assume — the effect that young workers have on aggregate unemployment. Surprisingly, he finds that a state's unemployment rate *falls* when its youth share rises. This negative correlation is not driven by the migration of young people to booming states. Using lagged birth rates to instrument for youth shares generates even larger unemployment declines. Shimer gives two interpretations to his findings. First, he concludes that firms want to locate in states with many young workers, because these workers are likely to be mismatched in their current jobs and accept other job offers. Second, the large number of vacancies posted by firms in "young" states lowers unemployment among all demographic groups.[3]

In this paper, I illustrate a pitfall in the use of state-level data for macroeconomic analysis, with the specific implication that the traditional model of demographic change in labor markets is not rejected after all. Shimer's sample period ends in 1996, but running his regressions through 2005 generates much weaker results. In most of the specifications I investigate, the absolute value of the youth-share coefficient falls by more than half when

[1] A good example of this approach is Aaronson et al. (2006), who study how the aging of the baby boom cohort would be expected to affect the national labor-force participation rate in the coming decades. Jaimovich and Siu (2007) investigate the effect of young workers on the volatility (not level) of economic activity across different countries.

[2] See also Bleakley and Fuhrer (1997) and Shimer (1998) on this point.

[3] In addition to the unemployment results, Shimer's paper includes evidence from wages and from manufacturing job creation and destruction rates that is also consistent with the predictions of the search-based model. Shimer stresses that his model is appropriate only for state-level unemployment (as opposed to national unemployment), because a constant cost of capital limits the amount of vacancy posting on the national level. Foote (2002) accepts Shimer's negative correlation as a fact, but argues that it is caused by a youth-induced housing boom, not by search considerations. I discuss in the conclusion why Foote's paper also suffers from a spatial-correlation problem.

using the updated data, in some cases by 70 to 90 percent. The reason, I argue below, is not because of structural changes in the economy that have rendered Shimer's search model less appropriate in recent years. Rather, large changes in the estimated youth-share effect should be expected, because the pre-1997 coefficients are not precisely estimated. As is typical in macroeconomic studies using regional or state-level panel data, Shimer assumed that the data from each state are independent draws from underlying distributions. In reality, state boundaries are often arbitrary political designations that divide nearly identical parts of the country. In some research designs, economic similarity across a state border is a good thing. In Card and Krueger's (1994) study of the effects of a change in New Jersey's minimum wage, Pennsylvania serves as a control state precisely because of the assumed similarity in the other economic shocks that affect the two states. However, in traditional panel regressions where both the outcome variable and the regressor of interest are functions of state-level economic or social climates, spatial correlation can lead to imprecise estimates and misleading standard errors.

The paper proceeds as follows: Section 2 illustrates how the estimated youth-share coefficient changes when Shimer's sample period is updated. The section also discusses the adjustments needed for the standard errors when both spatial and serial correlation are present. Importantly, for this macroeconomic problem, these adjustments are more complicated than simply clustering the covariance matrix by state and year simultaneously, as has been recommended in some microeconomic contexts with multi-way correlations.[4] The methods I use generate standard errors that are several times larger than the ones that Shimer reported, so that the pre-1997 estimates are no longer significant when the new methods are used. I also discuss why the methods I use may be imperfect, so that the larger standard errors I report may still be too small. In Section 3, I make some rough attempts to control for both serial and spatial correlation in the estimation procedure, not just in the calculation of the standard errors. Using data through 2005, these regressions generate estimates of the youth-share effect that are not only positive but also very close to what a mechanical model of demographic effects would imply. Unfortunately, the standard errors

[4] Recent work by Bertrand, Duflo, and Mullainathan (BDM, 2004) and Kézdi (2004) has highlighted the importance of accounting for serial correlation in panel data, following a line of research that goes back to Nickell (1981), Bhargava, Franzini, and Narendranathan (1982), and Solon (1984). Using simulations, BDM and Kézdi show that clustering the standard errors by state is usually adequate to account for serial correlation in state-level data, because the number of clusters (50) is relatively large. Cameron, Gelbach, and Miller (2006) and Thompson (2006) show that it is relatively simple to cluster in many dimensions simultaneously, which can be useful when both serial and spatial correlation are present. These papers are discussed in more detail below.

remain large, so precise inference is impossible. Section 4 concludes with the two main lessons of the paper. The specific lesson is that standard views on demographic change in labor markets are not refuted by U.S. state-level data. A more general implication is that macroeconomists should be wary of spatial correlation when testing theories with regional or sub-national panel data, especially when identification is achieved using instrumental variables.

2. Young Workers and Unemployment: 1973–2005

The basic regression in Shimer (2001) projects the state-level unemployment rate on the youth share and fixed effects for both state and year:

$$lnUR_{it} = \beta \; lnyshare_{it} + \phi_i + \phi_t + \epsilon_{it}, \tag{1}$$

where $i \in (1 \ldots N)$ indexes the state, $t \in (1 \ldots T)$ indexes the year, $lnUR$ is the natural log of the unemployment rate, and $lnyshare$ is the log of the share of the state's working-age population (ages 16-64) who are aged 16-24. If the youth share had only a mechanical effect on the overall unemployment rate, then the expected value of $\hat{\beta}$ is positive, in the neighborhood of .30. To see this, assume constant, age-specific unemployment rates for 16-24 year olds (UR^{young}) and 25-64 year olds UR^{old}. Denote the difference between these rates as $\widetilde{UR}$. Then the relationship between the levels (not logs) of the overall unemployment rate and the youth share is $UR = (yshare \cdot \widetilde{UR}) + UR^{old}$. Differentiating this expression with respect to the youth share and performing some algebra to obtain an elasticity gives

$$\frac{\Delta UR}{UR} = \left[\frac{\widetilde{UR}}{UR} \cdot yshare\right] \frac{\Delta yshare}{yshare}.$$

Using BLS data for both the population and the unemployment rate, the term in square brackets averages .29 from 1973 to 1996.[5] By contrast, Shimer's OLS estimate of β, using unbalanced panel data from 1970 to 1996, is a surprising -1.221, with a reported standard error of .160.[6]

Shimer then addresses two potential problems with this estimate. The first is that both unemployment rates and youth shares are positively serially correlated, so an OLS estimate

[5] Note that this calculation assumes that people stop working at 65. In the data, the fraction of the labor force corresponding to persons 65 and older never exceeds 3.5 percent during the sample period.

[6] This estimate is found in Table I, Panel A, Column 1 of Shimer (2001). The AR1-corrected estimate discussed in the next paragraph is found in the corresponding position in Panel B of the same table.

is likely to be inefficient (though consistent) and the standard errors will be biased down. Yet Shimer's AR1-corrected point estimate is almost identical to that from OLS: -1.219, with an standard error of .264. A second potential problem is endogeneity. Young people are likely to migrate to states with relatively low unemployment rates, causing a spurious negative correlation between low unemployment rates and high youth shares. To account for this possibility, Shimer uses lagged birth rates as an instrument for the youth share. This instrument is relevant, as Shimer shows that lagged birth rates account for a large majority of the variation in state-level youth shares. The IV is also plausibly exogenous, because current economic fluctuations can have no effect on birth rates 16–24 years in the past. Yet the youth-share coefficient becomes even more negative when using IV.[7]

Revisiting the unemployment-youth share relationship

Table I provides some new estimates of Shimer's regressions. Column 1 uses a sample that ends in 1996, as Shimer's did, while the sample for Column 2 ends in 2005.[8] Each of the four panels in Table I corresponds to a different estimation method. Panel A uses OLS, Panel B uses the lagged birth rate to instrument for the youth share, Panel C uses a Prais-Winsten AR1 correction, and Panel D uses both a Prais-Winsten correction and IV (where all variables, including the birth-rate instrument, are quasi-differenced by the Prais-Winsten procedure).[9] The most striking feature of Table I is the large change in the point estimates when the longer sample period is used. The absolute value of the OLS point estimate in Panel A drops by more than 70 percent, from -1.55 to -.42. The decline in the IV estimate in Panel B is an even steeper 90 percent, from -1.90 to -.19. The declines in the AR1-corrected estimates are not as severe, but the IV-AR1 estimate in Panel D still declines by more than half, from -1.68 to -.82.

[7] Using a restricted sample period (1978 to 1996) and an AR1 correction, Shimer's coefficient estimate is -1.466 under OLS and -1.807 under IV. See his Table IIa on p. 980.

[8] Column 1 is not meant to be a precise replication of Shimer (2001). My estimates differ slightly from his owing primarily to a recent benchmark revision of the state-level unemployment rates by the Bureau of Labor Statistics. In addition to slightly changing the unemployment data, this revision also allows a balanced sample of states, starting in 1973, which I exploit throughout this paper. I also exclude Alaska, Hawaii, and the District of Columbia and use updated measures of the youth share from SEER (2005); these data are described in Ingram et al. (2003). In practice, these differences do not have material effects when similar sample periods are used.

[9] Nickell (1981) points out that there is a "short-T" bias when estimating dynamic, fixed-effects models. Accordingly, for all the AR models in this paper, I use estimates of the AR parameters that are corrected by the method of Hansen (forthcoming *a*).

There are two potential explanations for these changes. One is that the earlier estimates were close to the truth, but that structural changes in the way that workers match to jobs have rendered Shimer's model less appropriate in recent years. The other possibility is that changes in the coefficients should be expected, because the pre-1997 coefficients are not precisely estimated. In this scenario, the addition of more recent data brings the estimates closer to the truth. The remainder of this section presents evidence in favor of the latter explanation, by showing how accounting for spatial correlation increases the standard errors and undermines the statistical significance of the pre-1997 coefficients.

Shimer (2001) reports unadjusted standard errors, which turn out to be nearly identical to Huber-White "robust" errors. These errors account for heteroskedasticity in individual residuals, but not for correlations across years and states:

$$V_{White} = (X'X)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} x'_{it}\hat{\epsilon}^2_{it}x_{it}\right)(X'X)^{-1}, \tag{2}$$

where x_{it} is the $(1 \times K)$ vector of regressors for state i at time t, $\hat{\epsilon}$ is an estimated residual, and X is the full data matrix.[10] These errors are reported in the first row of standard errors of each panel. For the pre-1997 sample, they imply very large t-statistics in all cases.

Under the assumption of serial but no spatial correlation, these errors are valid in Panels C and D when the serial correlation follows an AR1 process. Because the estimators in Panels A and B include no correction for serial correlation, they require standard errors that are clustered by state when serial correlation is present. Let ϵ denote the full $(NT \times 1)$ vector of errors and let ϵ_i denote the $(T \times 1)$ vector corresponding to state i. If the data are sorted by state, so that $\epsilon' = [\epsilon'_1 \dots \epsilon'_i \dots \epsilon'_N]$, then state-clustering assumes that the $NT \times NT$ matrix $\Omega = E(\epsilon\epsilon')$ is block-diagonal, so that correlations exist only among residuals corresponding to the same state. We can write

$$\Omega = E(\epsilon\epsilon') = \begin{pmatrix} \Sigma_{11} & & & & 0 \\ & \ddots & & & \\ & & \Sigma_{ii} & & \\ & & & \ddots & \\ 0 & & & & \Sigma_{NN} \end{pmatrix},$$

[10] Throughout this paper, I define covariance estimators using OLS formulas to provide intuition. Under IV, the center matrices in these formulas would obviously involve the products of the instruments and the error terms, not the regressors and the error terms.

where $\Sigma_{ii} = E[\epsilon_i \epsilon_i']$. The state-clustered covariance matrix is then

$$V_{state} = (X'X)^{-1} \left(\sum_{i=1}^{N} x_i' \hat{\epsilon}_i \hat{\epsilon}_i' x_i \right) (X'X)^{-1}, \tag{3}$$

where x_i is the $(T \times K)$ matrix of regressors for state i. These errors are reported in the second row of each panel. As expected, the use of state-clustered errors has the largest effects in Panels A and B, where no AR1 corrections are performed. In all panels, however, state-clustering generates t-statistics that remain significant at the 5 percent level in the pre-1997 sample.

The effect of spatial correlation

The third set of standard errors in Table I are clustered by year rather than state. They therefore account for spatial correlation, but not for serial correlation. The corresponding covariance matrix is figured analogously to the state-clustered matrix, but sorts the data by year rather than state: $\epsilon' = [\epsilon_1' \ldots \epsilon_t' \ldots \epsilon_T']$. This gives

$$V_{year} = (X'X)^{-1} \left(\sum_{t=1}^{T} x_t' \hat{\epsilon}_t \hat{\epsilon}_t' x_t \right) (X'X)^{-1}. \tag{4}$$

Table I shows that these year-clustered errors are always larger than the state-clustered ones, indicating that in this context, spatial correlation may be more damaging for inference than serial correlation. The effect of year-clustering is biggest in the AR1-corrected panels, where the errors more than double, to .73 in Panel C (OLS-AR1) and 1.02 in Panel D (IV-AR1).

To get a sense of why the year-clustered errors are so large, Figure 1 presents choropleth maps of spatial correlation in unemployment and youth shares among U.S. states for a sample year (1985). In order to illustrate the variation that identifies a youth-share coefficient in a regression with state and year fixed effects, both the unemployment rates and the youth shares are deviates from their respective state and year means. The maps show that the relative values of both variables for one state are generally close to corresponding values in nearby states. For example, Panel A shows that when relative unemployment rates were low throughout New England in 1985, the "Massachusetts Miracle" was experienced in Rhode Island and Connecticut, too. Similarly, Panel B shows that all New England states, not just Massachusetts, experienced relatively high youth shares at the same time.

The maps illustrate spatial correlations for only one year. Cross-year comparisons in the intensity of spatial correlation are facilitated by parameterizing this correlation with the familiar spatial autoregressive model. For unemployment in state i in a single year, this model is

$$lnUR_i = \lambda W_N \mathbf{lnUR} + v_i,$$

where W_N is a known $N \times N$ spatial weighting matrix in which the (i,j)th element indicates the "closeness" of state i to state j, and $\mathbf{lnUR}$ represents the $N \times 1$ vector of unemployment rates for all states in the given year. The scalar λ measures the intensity of spatial correlation, and v_i is a residual. Estimates of λ depend on the distance metric assumed for the matrix W_N. A common choice for W_N is a first-order contiguity matrix, where the (i,j)th element of W_N equals one if state i and state j share a common border and zero otherwise. Also, it is common to row-standardize the weighting matrix, so that the sum of each row equals 1. When using a contiguity matrix for W_N, row-standardization allows an interpretation of (say) $\lambda = 0.5$ to mean that the unemployment rate for state i equals one-half of the average unemployment rate of the states that surround it, plus an idiosyncratic error term v_{it}.

Figure 2 presents yearly estimates of λ for both unemployment and youth shares from 1973 to 2005. Also graphed are 95-percent confidence intervals. These estimates are generated by 33 separate maximum-likelihood regressions, one for each year of the data.[11] For most years, the estimates of λ are significantly different from zero. In the mid- to late-1980s, spatial correlation in unemployment and youth shares was especially large, with point estimates of λ exceeding 0.5 for both variables.[12]

High degrees of spatial correlation imply that clustering the covariance matrix by year is a good idea, but this does not mean that we cannot cluster the errors by state as well. Cameron, Gelbach, and Miller (2006) and Thompson (2006) point out that a multi-way clustered covariance matrix can be constructed by adding the two clustered covariance matrices together, then subtracting the relevant White matrix to avoid double counting.[13]

[11] As Anselin (1988) explains, this model must be estimated by maximum likelihood rather than OLS. By pre-whitening the data before running the individual MLE regressions, I ignore the incidental parameters problem that arises in maximum-likelihood regressions with fixed effects.

[12] Results were similar using an inverse-distance weighting matrix with a cutoff of 500 miles, rather than the first-order contiguity matrix. The spatial regressions make extensive use of the MATLAB code available on James LeSage's website (www.spatial-econometrics.com).

[13] The subtraction of the White matrix is required because both the V_{state} and V_{year} matrices involve

In our case, the multi-way clustered matrix is

$$V_{state-year} = V_{state} + V_{year} - V_{White}. \tag{5}$$

Standard errors generated by this method appear in the fourth rows of each panel in Table I. As expected, adding V_{state} to V_{year} has the biggest effect in Panels A and B, where no AR1 corrections are performed.

A goal of Table I is to show that accounting for wider covariance patterns among residuals undermines the statistical significance of the youth-share coefficients in the pre-1997 data. To some extent this goal is accomplished. The estimates in Panel D of Column 1 (IV-AR1) are no longer significant at the 10-percent level. However, even clustering by state and year generates t-statistics that are significant at the 10-percent level or higher in all the other panels. Moreover, the estimates remain strongly significant in A and B.

It turns out that the t-statistics in Table I are so resilient because even $V_{state-year}$ does not account for all the problematic correlations in these data. In the $V_{state-year}$ formula above, the presence of the V_{state} matrix accounts for correlations belonging to the same state, while that of the V_{year} matrix does the same for errors belonging to the same year. But nothing in $V_{state-year}$ accounts for the correlation between a state's error in one year and a nearby state's error in the following year. This type of correlation is likely to exist in macroeconomic panel data. A boom that generates low unemployment in one state and year may contribute to favorable economic conditions in neighboring states in following years. If this type of variation is also found in the right-hand-side variables, then $V_{state-year}$ will not provide the appropriate standard errors. What is needed is an estimate of the covariance matrix that is general enough to account for a wide variety of error correlations, even those that span different states and years.

Driscoll and Kraay (1998)

There are currently at least three approaches to obtaining a consistent covariance matrix for more general correlation structures. A paper by Driscoll and Kraay (DK, 1998) provides not only a candidate solution but also a useful framework for thinking about the problem. DK point out that the panel-data inference problem with general serial patterns and spatial correlation can be thought of as a time-series problem in the cross-sectional *means* of

the term $x'_{it}\epsilon_{it}\epsilon'_{it}x_{it}$.

the products of the regressors and error terms. If these $(K \times 1)$ products are denoted $h_{it} = x'_{it}e_{it}$, then the relevant cross-sectional mean for period t is $\overline{h}_t = \frac{1}{N}\sum_{i=1}^{N} h_{it}$. The time-series behavior of these means must be accounted for when constructing the covariance matrix, and DK provide the specific conditions where the standard Newey-West technique can be applied.

Though DK do not make the connection, their estimator has a cluster interpretation. Working with the cross-sectional means of h_{it} is equivalent to clustering by year, and using the Newey-West method to account for serial correlation in $\overline{h}_t$ allows for correlations that span different states and years. Denote

$$V_{year,l} = (X'X)^{-1}\left(\sum_{t=l+1}^{T} x'_t\hat{\epsilon}_t\hat{\epsilon}'_{t-l}x_{t-l}\right)(X'X)^{-1}, \tag{6}$$

where the simple year-clustered matrix corresponds to $l = 0$.[14] I show in the appendix that the DK estimator can be written

$$V_{DK(m)} = V_{year} + \sum_{l=1}^{m} w(l,m)\ \left(V_{year,l} + V'_{year,l}\right), \tag{7}$$

where m is a maximal lag length over which serial correlation is allowed. DK's method reduces to year-clustered standard errors $(m = 0)$ when spatial correlation is an issue but serial correlation is not. When serial correlation is also present, DK's method allows for the off-diagonal blocks of the Ω matrix to be non-zero up to a maximal lag length m, smoothing these estimated correlations with the linear weights w. As in the standard Newey-West setup, m is assumed to grow with T, so the procedure is consistent for a variety of correlation structures as $T \to \infty$.

Table II includes standard errors for equation (1) using DK's method. (The parameter estimates and the state-year clustered errors in this table are reprinted from Table I for ease of comparison.) The second and third rows of Table II are the DK errors with the maximal lag length m set to one and three years, respectively. In general, the DK errors are much larger than the simultaneous state-year cluster when $m = 3$. In Column 1 of Panel A (OLS), the DK(1) standard error (.66) is only slightly larger than the state-year

[14] One should think of the data as sorted by year here, so that $\epsilon' = [\epsilon'_1 \ldots \epsilon'_t \ldots \epsilon'_T]$. As the appendix illustrates, the cluster interpretation of the DK estimator follows from noting that $\overline{h}_t = \frac{1}{N}\sum_{i=1}^{N} x'_{it}\epsilon_{it} = \frac{1}{N}x'_t\epsilon_t$.

clustered error (.61), so that the corresponding coefficient estimate is still significant at the 5-percent level. But the DK(3) error (.80) is more than 30 percent larger than the state-year error, so that the significance level falls to the 10-percent level. In Panel B (IV), the error rises from .82 using the state-year cluster to 1.09 using DK(3). In the last two panels, the DK errors rise from .75 to .89 (OLS-AR1), and from 1.02 to 1.42 (IV-AR1).

DellaVigna and Pollet (2007)

The DK estimator is attractive for several reasons. It is not only easy to calculate[15] but also fully non-parametric, so it requires no specific assumptions for the serial or spatial processes. But this flexibility comes at a price, because the standard Newey-West estimator is likely to underestimate standard errors for persistent series in short samples (Andrews 1991). This drawback is likely to be substantial in panel data, where T is typically smaller than in pure time-series applications. A second covariance estimator, due to DellaVigna and Pollet (DVP, 2007), imposes a parametric assumption on the serial correlation in hopes of obtaining a better estimate. Using DK's notation, DVP essentially assume that $\overline{h}_t$ follows an AR1 process: $\overline{h}_t = \rho\overline{h}_{t-1} + \nu_t$. An estimate of ρ is easily obtained by regressing each of the K elements of $\overline{h}_t$ on once-lagged values. DVP show that with an estimate of $\hat{\rho}$ in hand, the covariance matrix has a simple form:

$$V_{DVP} = \left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right) V_{year}. \tag{8}$$

The DVP errors are shown in the fourth rows of Table II.[16] They are sometimes larger and sometimes smaller than the DK estimates. The DVP errors are larger in Panels A and B (no AR1 corrections). Because these regressions leave a great deal of serial correlation in the residuals, the parametric AR1 assumption in the DVP errors may do a better job of capturing this serial correlation than the Newey-West approach of DK.[17] By contrast, in

[15] The DK method is now available as the `xtscc` add-on to the Stata statistical package; see Hoechle (2007). My implementation of the DK errors differs slightly from Hoechle and DK, because I normalize the covariance matrix by $\left(\frac{NT-1}{NT-K}\right)\left(\frac{T}{T-1}\right)$. This is Stata's small-sample normalization for a year-clustered covariance matrix and quite close to the suggested normalization for the cluster in Hansen (forthcoming *b*) in this context: $\left(\frac{T}{T-1}\right)$.

[16] See the appendix for specific details on how the $\hat{\rho}$'s used to construct these errors were estimated. The method I employ differs slightly from the one in DVP's original paper.

[17] In Column 1, the estimated value for ρ in both Panel A (OLS) and Panel B (IV) is .79. This high degree of correlation is likely to cause problems for a Newey-West estimator in a sample with short T.

the AR1-corrected regressions (Panels C and D), the DVP errors are smaller than the DK errors. In these regressions, a smaller estimate of ρ used to construct the DVP errors is to be expected, because these regressions purge AR1 correlation from the data beforehand.[18] The fact that DK errors are larger than the DVP errors in these bottom panels suggests that additional serial correlation remains in the data even after the AR1 corrections, so that the DVP errors are too small.

Thompson (2006)

Both the DVP and DK approaches account for serial correlation in h_{it} within an individual state only to the extent that this correlation affects the relationships among the cross-sectional means $\overline{h}_t$. However, a state-clustered matrix is likely to account well for correlation in h_{it} if the number of clusters (states) is large enough. A third covariance estimator, suggested in Thompson (2006), takes advantage of this fact by adapting the multi-way cluster for correlations across different states and years. His estimator is

$$V_{Thompson(m)} = V_{state} + V_{year} - V_{White} + \sum_{l=1}^{m} V_{year,l} + V'_{year,l} - V_{White,l} - V'_{White,l}, \quad (9)$$

where, as with the $V_{state-year}$ matrix, the subtraction of the V_{White} and $V_{White,l}$ matrices is required to avoid double counting.[19] In this expression, the presence of V_{state} controls for within-state correlation. Correlations that span different states and years are assumed to follow an MA(m) process, dying off after m periods. There is no assumption (as in DK) that m grows with T, in an attempt to capture arbitrary serial correlation processes.

The Thompson errors are presented in the last two rows of each panel of Table II, with $m = 1$ and $m = 3$, respectively. Not surprisingly, the errors tend to be larger than the DK errors. Unlike DK's, Thompson's errors include no smoothing weights on the $V_{year,l}$ matrices. Moreover, the within-state cluster V_{state} in the Thompson formula is likely to do a better job of accounting for purely within-state serial correlation than the Newey-West approach of DK. In Panels C and D, the Thompson errors are also larger than the DVP errors. As noted in the previous paragraph, the AR1 corrections in these panels reduce the effect of using DVP's method, but Thompson's method is able to capture more general serial correlation patterns than AR1. In any case, using Thompson's method with $m = 3$,

[18] The estimated values of ρ using 1973–1996 data are only .08 and .10 in Panels C and D, respectively.

[19] Note that $V_{White,l} = (X'X)^{-1}(\sum_{i=1}^{N}\sum_{t=1}^{T-l} x'_{it}\hat{\epsilon}_{it}\hat{\epsilon}_{i,t-l}x_{i,t-l})(X'X)^{-1}$.

none of the point estimates in the original sample period remain significant in any panel of the table.

But even the Thompson(3) errors may still be too small if the cross-state correlations do not die off after three years. If so, then we would have to increase m, but here again we run into a short-T problem, as we did with DK. Figure 3 graphs the value of the Thompson errors for values of m ranging from one to seven years. Each panel corresponds to one of the four estimation methods we have examined so far (OLS, IV, OLS-AR1 or IV-AR1).[20] The darker lines in each of the four panels correspond to the errors from the shorter sample period (1973–1996), where $T = 24$. The lighter lines correspond to Thompson errors from the longer sample (1973–2005), where $T = 33$. Consider first Panels A (OLS) and B (IV) in the top row, which correspond to the regressions without AR1 corrections. The darker lines are hump-shaped, indicating that in the shorter sample, the Thompson errors actually get smaller when m exceeds three or four years. This is probably due to the poor statistical properties that result when long correlations are estimated with a short time series. Note that the gray lines in these same panels level off after three or four periods, rather than decline. The larger T used for the gray lines no doubt does a better job of capturing cross-state correlations at longer lags. It is hard to know whether using a T that is even greater than 33 would cause the standard errors to continue to rise for $m > 3$, indicating that correlations at lags greater than m should be included in the Thompson procedure. The story is similar in Panels C and D of the bottom row, which correspond to the AR1-corrected regressions. Here, the hump shape in the short-sample errors is less pronounced; after rising when m changes from 0 to one, the short-sample errors essentially flatten out. However, the longer-sample errors continue to rise smoothly as m increases from zero to five or six. The long-sample pattern suggests that cross-state correlations exist at longer values of m, but the short-sample results suggest that these correlations cannot be estimated well with a T of only 24.

This bottom-row pattern shows how even the most careful researcher could be tripped up by various correlations in the data. Consider a researcher who comes to these data knowing that both spatial and serial correlation are present. She might use an AR1 correction in her regression to purge serial correlation in the estimation process. When calculating the standard errors, she is concerned about serial correlation that remains after this AR correction, as well as spatial correlation, causing her to cluster by both state and year,

[20] Note that setting $m = 0$ corresponds to $V_{state-year}$.

using $V_{state-year}$. Realizing that cross-state correlations may exist at lag lengths greater than $m = 0$, she checks to see whether the Thompson errors at various values of m are much larger than estimated errors from the state-year cluster (where $m = 0$.) If only the shorter sample were available, this method would imply that choosing $m = 1$ essentially "catches" all of the cross-state correlation. But this pattern could also result from using a sample where T is too short to measure cross-state correlations at long lags, so that the resulting standard errors are deceptively small.

3. New Estimates with Additional Regressors

Given the difficulty of accounting for both serial and spatial correlation in these data, it is worthwhile to control for as much of it as possible in the regression, using additional covariates. The added regressors should soak up spatial correlation in the errors (to improve precision) but should also be exogenous with respect to innovations in state-level unemployment rates (to preserve identification). One possibility is the "shift-share" measure of state-level labor demand that was originally suggested by Timothy Bartik (1991). Bartik's goal was to create a measure of state-level labor-demand shocks that would not be contaminated by innovations in state-level labor supply. His measure is a weighted average of national, industry-level growth rates, using state-specific industry weights from a given base year. With fixed state-level weights, time-series variation in this variable is generated solely by fluctuations in industry growth rates on the national level. The variable is therefore exogenous to innovations in state-level youth shares. The inclusion of the Bartik variable in equation (1) will soak up some spatial correlation if the industry weights are spatially correlated (for example, if most U.S. automobile production were concentrated in the Upper Midwest). Of course, spatial correlation is likely determined by a host of factors besides industry mix, so this method is likely to leave much of the spatial correlation unaddressed.

An additional, "brute force" way of accounting for spatial correlation is simply to interact the yearly dummies with dummies corresponding to given geographical areas, such as the country's four Census regions or its nine Census divisions.[21] While easy to implement, this method is also imperfect. First, the interactions assume that the spatial effect is

[21] The Census divisions are New England (CT, ME, MA, NH, RI, and VT), Middle Atlantic (NJ, NY, and PA), East North Central (IN, IL, MI, OH, and WI), West North Central (IA, KS, MN, MO, NE, ND, and SD), South Atlantic (DE, FL, GA, MD, NC, SC, VA and WV), East South Central (AL, KY, MS, and TN), West South Central (AK, LA, OK, TX), Mountain (AZ, CO, ID, NM, MT, UT, NV, and WY), and Pacific (CA, OR, and WA). The Northeast Region consists of the New England and the Middle Atlantic divisions, the Midwest Region comprises the East North Central and West North Central divisions, the

constant within each particular region for a given year. The larger the region, the less appropriate this assumption will be. But choosing smaller regions worsens a second problem: bordering states may be assigned to different regions, even though the true correlation in their errors is large.

Table III presents estimates of the unemployment equation using the additional variables and the longer sample period (1973–2005). When all variables are included, the regression is

$$lnUR_{it} = \beta \; lnyshare_{it} + \sum_{p=0}^{3} \gamma_p \; Bartik_{i,t-p} + \phi_i + \phi_{rt} + \epsilon_{it}, \qquad (1')$$

where $Bartik_{it}$ denotes the shift-share labor-demand variable and ϕ_{rt} denotes either region-year or division-year interactions. The first column of Table III replicates the regression from Tables I and II without the additional variables and is included for comparison. The Bartik variable and three lags are introduced in Column 2. These regressors have modest effects on the point estimates, moving them from -.42 to -.29 (OLS in Panel A) and from -.19 to -.21 (IV in Panel B). Interestingly, the inclusion of these variables reduces the year-clustered and Thompson standard errors by appreciable margins, suggesting that the variables are indeed adding precision to the regression. Column 3 adds (4 - 1) × (33 - 1) = 96 region-year interactions. These variables generally reduce the year-clustered and Thompson standard errors further. More importantly, however, the inclusion of these variables moves the IV point estimate to almost exactly what we would expect with a mechanical model (.30). Unfortunately, the standard errors remain too large to take a precise stand on this coefficient. Column 4 replaces the region-year interactions with (9 - 1) × (33 - 1) = 256 division-year interactions. The year-clustered and Thompson errors decline further under OLS but are generally unaffected under IV. The IV point estimate rises slightly, to .35. Finally, Columns 5 and 6 impose AR1 and AR2 corrections, respectively. These corrections reduce IV point estimates somewhat, to .21 and .30, but the estimates remain positive and in the neighborhood of what we would expect with a standard model. The AR corrections have a much larger effect on the OLS estimates, making them more negative (-.51 and -.58).

All in all, the last few columns of Table III are consistent with traditional approaches to measuring demographic change in labor markets. The IV estimates in the lower panel

South Region is the South Atlantic, East South Central, and West South Central divisions, and the West Region is made up of the Mountain and Pacific divisions.

suggest that exogenous increases in a state's youth share will raise its unemployment rate by about as much as a mechanical model would predict. By contrast, the OLS point estimates are negative, suggesting that young people do indeed move to states with low unemployment rates, causing a spurious negative correlation between youth shares and unemployment on the state level. Indeed, the OLS estimate in the full model of Columns 5 and 6 is even more negative than in the initial model of Column 1.[22] Unfortunately, these point estimates are only suggestive, because the standard errors in Table III remain very large.[23] In short, there is simply not enough spatially independent variation in unemployment and youth-shares across U.S. states to test formal theories about demographic effects on labor markets with these data, even with the longer sample period and the inclusion of regressors to absorb spatial correlation in the errors.[24]

4. Conclusions

This paper makes two main points. The first is that spatial correlation has a profound effect on the precision of these estimates, so that evidence against mechanical effects of demographic change in labor markets is not as strong as it first appears. The cleanest way to quantify this effect of spatial correlation is to compare spatially corrected errors in well-identified (that is, IV) regressions to those that would be valid if no spatial correlation were present. In Panel B of Table II, (IV), the Thompson(3) error of 1.26 is about 2.5 times as large as the corresponding state-clustered error in Table I (.53). In Panel D of Table II (IV-AR1), the Thompson(3) error is 1.52 in the 1973–1996 data, more than three times

[22] One can test directly for the youth-migration effect by regressing the youth share on unemployment and the lagged birth rates, using the Bartik variables to instrument for unemployment. Doing so generates significantly negative coefficients on the unemployment rate, suggesting that young people are more likely to move to low-unemployment states than are older people (so that the youth share rises when unemployment falls). The estimates coefficients, however, are small and account for little of the variation in youth shares.

[23] Estimating the regressions of Table III on pre-1997 data also results in insignificant coefficients, but all of the IV estimates remain negative. Specifically, using division-year dummies and the Bartik variables in the IV-AR2 regression of Column 6 generates a point estimate of -.45, as compared with the corresponding estimate of .30 in Table III. However, this estimate is still much different from the IV estimates of -1.90 and -1.68 from Table I, suggesting that the attempts to soak up some spatial correlation are moving the estimate in the right direction.

[24] In recent years a number of authors have devised dynamic panel estimators that account for spatial correlation using a formal weighting matrix rather than geographic interactions. See, for example, Elhorst (2003a, 2003b, 2003c), Yu, DeJong, and Lee (2006), Lee and Yu (2007) and Su and Yang (2007). These estimators may well prove more efficient than the regressions I estimate here. A disadvantage of using these more formal spatial estimators is that they assume that the intensity of spatial correlation is constant over time. In our data, that assumption may be contradicted by the top panel of Figure 2, which shows that the estimate of λ fluctuates from .87 (in 1988) to .02 (in 2002).

the size of the Huber-White standard error in the first row of Panel D in Table I (.44). Moreover, the spatially corrected errors are also multiples of the first-order youth-share effect that we should expect if the mechanical model were true, and they leave none of the parameter estimates significant.

A second, more general lesson is that macroeconomists should be careful when using state-level or regional panel data, because cross-sectional units may not generate adequate independent variation, and accounting for cross-state correlations at long lags is difficult when T is short. In recent years, the use of cluster estimators has grown as applied researchers have come to discover their key desirable property: as long as there are enough clusters, the method can control for arbitrary correlation structures.[25] In this context, however, we have seen that while clustering helps with some issues, the problem of correlations that span different states and years does not have a simple "cluster fix." Assuming that cross-correlations die off after three years in our data was enough to render pre-1997 estimates insignificant using Thompson's method, but even these standard errors may be too small, given the small number of years available to estimate correlations at longer lags.

The use of instrumental variables adds an interesting wrinkle to this issue. IV approaches obviate the need for formal models of the error term, as long as the instrument is both exogenous and relevant. But these conditions imply only that with infinite data, the sampling distribution of the IV estimator collapses to a spike at the coefficient's true value. With finite data, there will be *some* variance to this sampling distribution, which must be estimated. During the past decade, a great deal of research has focused on the problem of estimating this distribution when instruments are only weakly correlated with the regressors. Our application does not have a weak-instruments problem, because lagged birth rates are strongly correlated with subsequent youth shares. Our problem is that using *only* youth shares as a regressor leaves a great deal of other influences in the error term. These influences are both spatially and serially correlated, as is the regressor of interest. As a result, the researcher may be forced to make (and test) some parametric assumptions about the nature of these correlations to insure adequate inference, even if the instrument itself fulfills all of the usual requirements.

Even more problematic is the use of data from adjoining cross-sectional units to construct instruments. Indeed, if Shimer (2001) commits an econometric misdemeanor by

[25] See Hansen (forthcoming *b*) for a discussion of how the cluster estimator might be useful even with a "small" number of clusters.

ignoring spatial correlation, then Foote (2002) commits a felony. Foote accepts the negative coefficients in Shimer's original paper as informative, but he argues that they stem from youth-induced housing booms, not from search-model considerations. To support this claim, Foote adds a measure of state-level housing construction to equation (1), to see whether it knocks out the youth-share coefficient. Because construction is endogenous with respect to the unemployment rate, he needs another instrument, which he defines (for each state) as the weighted average of lagged birth rates in "nearby" states.[26] The significance of the youth-share coefficient is sharply reduced when the construction measure is entered, but the additional, geographically determined instrument is seriously compromised if the errors are spatially correlated.

[26] The additional instrument will be relevant if young people in nearby states also put pressure on the home state's housing market. "Nearness" of state j to state i is defined as the share of in-migrants to state i that is accounted for by state j, not as the physical distance between states i and j.

5. Appendix

Driscoll-Kraay (1998) as a cluster estimator

Consider the sampling error of equation (1),

$$(\hat{\beta} - \beta) = \left(\frac{X'X}{NT}\right)^{-1} \left(\frac{X'\epsilon}{NT}\right), \tag{10}$$

where the fixed effects ϕ_i and ϕ_t are omitted for clarity. Define the product of regressors and residuals for state i at time t as $h_{it} = x'_{it}\epsilon_{it}$. This gives

$$(\hat{\beta} - \beta) = \left(\frac{X'X}{NT}\right)^{-1} \left(\frac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T} h_{it}}{NT}\right).$$

DK collapse the covariance-estimation problem into the time-series dimension by working with the cross-sectional *means* of h_{it}, denoting $\frac{1}{N}\Sigma_{i=1}^{N} h_{it} = \overline{h}_t$. Making this substitution into (10) and multiplying both sides by $\sqrt{T}$ gives

$$\sqrt{T}(\hat{\beta} - \beta) = \left(\frac{X'X}{NT}\right)^{-1} \cdot \sqrt{T}\left(\frac{\Sigma_{t=1}^{T}\overline{h}_t}{T}\right). \tag{11}$$

This expression involves $\sqrt{T}$ times the average (over T) of $\overline{h}_t$, so under standard regularity conditions, the limiting distribution of this vector will be mean 0 with variance $V_T = Q^{-1}S_T Q^{-1}$, where Q is the probability limit of $(\frac{X'X}{NT})$ and

$$S_T = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E(\overline{h}_t \overline{h}_s{}'). \tag{12}$$

In the absence of serial correlation, $E(\overline{h}_t \overline{h}_s{}')$ will equal 0 for all $t \neq s$. The matrix S_T then simplifies to $S_T = \frac{1}{T}\Sigma_{t=1}^{T} E(\overline{h}_t \overline{h}_t{}')$. When serial correlation is present, then S_T does not simplify. Specifically, if $\Gamma_l = E(\overline{h}_t \overline{h}_{t-l}{}')$, then

$$S_T = \Gamma_0 + \sum_{l=1}^{\infty} [\Gamma_l + \Gamma'_l].$$

DK provide the general restrictions on $\overline{h}_t$ where the assumptions in Newey and West (1987) hold, so that S_T can be estimated with

$$\widehat{S}_T = \widehat{\Gamma}_0 + \sum_{l=1}^{m} w(l, m)[\widehat{\Gamma}_l + \widehat{\Gamma}'_l],$$

with $\widehat{\Gamma}_l = \frac{1}{T}\sum_{t=l+1}^{T} \overline{h}_t \overline{h}_{t-l}{}'$ and $w(l,m) = 1 - \frac{l}{m+1}$ defining linear smoothing weights for covariances at lag l and maximal bandwidth m. The estimated asymptotic variance of (11) is then $\widehat{V}_T = \widehat{Q}^{-1}\widehat{S}_T\widehat{Q}^{-1}$.

To see the relationship of this estimator with the standard year-cluster estimator, note that $\overline{h}_t = \frac{1}{N}\sum_{i=1}^{N} x_{it}'\epsilon_{it} = \frac{1}{N}x_t'\epsilon_t$, so that

$$\begin{aligned}
\widehat{\Gamma}_l &= \frac{1}{T}\sum_{t=l+1}^{T} \overline{h}_t \overline{h}_{t-l}{}' \\
&= \frac{1}{T}\sum_{t=l+1}^{T} \left(\frac{1}{N}x_t'\epsilon_t\right)\left(\frac{1}{N}x_{t-l}'\epsilon_{t-l}\right)' \\
&= \frac{1}{T}\frac{1}{N^2}\sum_{t=l+1}^{T} x_t'\epsilon_t\epsilon_{t-l}'x_{t-l}.
\end{aligned}$$

Applying this result to the estimated covariance matrix for β, $V_{DK(m)} = \frac{\widehat{V}_T}{T}$, we can write

$$\begin{aligned}
\frac{1}{T}\widehat{V}_T &= \frac{1}{T}\widehat{Q}^{-1}\{\widehat{S}_T\}\widehat{Q}^{-1} \\
&= \frac{1}{T}\left(\frac{X'X}{NT}\right)^{-1} \\
&\quad \left\{\frac{1}{T}\frac{1}{N^2}\left[\sum_{t=1}^{T} x_t'\hat{\epsilon}_t\hat{\epsilon}_t'x_t + \sum_{l=1}^{m} w(l,m)\left(\sum_{t=l+1}^{T} x_t'\hat{\epsilon}_t\hat{\epsilon}_{t-l}'x_{t-l} + \sum_{t=1}^{T-l} x_{t-l}'\hat{\epsilon}_{t-l}\hat{\epsilon}_t'x_t\right)\right]\right\} \\
&\quad \left(\frac{X'X}{NT}\right)^{-1} \\
&= (X'X)^{-1}\left\{\left[\sum_{t=1}^{T} x_t'\hat{\epsilon}_t\hat{\epsilon}_t'x_t + \sum_{l=1}^{m} w(l,m)\left(\sum_{t=l+1}^{T} x_t'\hat{\epsilon}_t\hat{\epsilon}_{t-l}'x_{t-l} + \sum_{t=1}^{T-l} x_{t-l}'\hat{\epsilon}_{t-l}\hat{\epsilon}_t'x_t\right)\right]\right\}(X'X)^{-1}.
\end{aligned}$$

It is easy to see that this expression reduces to

$$V_{DK(m)} = V_{year} + \sum_{l=1}^{m} w(l,m)\ \left(V_{year,l} + V_{year,l}'\right)$$

when we define

$$V_{year} = (X'X)^{-1}\left(\sum_{t=1}^{T} x_t'\hat{\epsilon}_t\hat{\epsilon}_t'x_t\right)(X'X)^{-1}$$

and

$$V_{year,l} = (X'X)^{-1}\left(\sum_{t=l+1}^{T} x_t'\hat{\epsilon}_t\hat{\epsilon}_{t-l}'x_{t-l}\right)(X'X)^{-1}$$

as we do in equations (4) and (6) in the text.

Estimating the AR1 parameter in DellaVigna and Pollet (2007)

As noted in the text, the DellaVigna-Pollet standard errors in Table II are generated by scaling up the year-clustered covariance estimator V_{year} by $\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right)$, where $\hat{\rho}$ is estimated from the regression:

$$\overline{h}_t = \rho \overline{h}_{t-1} + \nu_t. \tag{13}$$

With T years in the data and K elements of $\overline{h}_t$, this AR1 regression will involve $K(T-1)$ observations. This regression is slightly different than one that DVP use for the application in their paper. They recover ρ with a pooled regression of each element of h_{it} (not $\overline{h}_t$) on its corresponding lagged value:

$$h_{it} = \rho h_{i,t-1} + \nu_{it}, \tag{14}$$

which will involve $NK(T-1)$ observations. However, the assumptions they use to derive the simple form of their estimator would suggest that using equation (13) is also appropriate. Specifically, DVP assume that

$$E\left[\left(\sum_{i=1}^{N} h_{i,t-p}\right)'\left(\sum_{i=1}^{N} \nu_{it}\right)\right] = 0$$

for all $p > 0$. This orthogonality condition allows ρ to be recovered with equation (13) above, since it concerns the behavior of the *sums* of h_{it} and ν_{it} over i (that is, $N\overline{h}_s$ and $N\nu_t$), not the h_{it}s and ν_{it}s themselves. From a theoretical standpoint, there is also an advantage to using equation (13) rather than equation (14) to estimate ρ. Aggregating $h_{it} = x_{it}'\epsilon_{it}$ across states before running the regression allows any correlations that span different states and years to directly inform the estimate of ρ.[27] As a practical matter for this paper, using equation (13) to recover ρ tends to generate larger values of $\hat{\rho}$, and therefore larger standard errors, than using equation (14).

[27] Consider a correlation between Michigan's value of h_{it} in 1979 and Indiana's value in 1978. This correlation can affect the estimate of ρ in equation (13), because Michigan's h_{it} contributes to $\overline{h}_t$ while Indiana's contributes to $\overline{h}_{t-1}$. This is not the case in equation (14), because both right-hand-side and left-hand-side variables correspond to the same state.

References

Aaronson, Stephanie; Bruce Fallick; Andrew Figura; Jonathan Pingle; and William Wascher (2006). "The Recent Decline in the Labor Force Participation Rate and Its Implications for Potential Labor Supply," *Brookings Papers on Economic Activity*, 2006:1, pp. 69-134.

Andrews, Donald W. K. (1991). "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59:3, pp. 817-58.

Anselin, Luc (1988). *Spatial Econometrics: Methods and Models* (Dordrecht: Kluwer).

Bhargava, A.; L. Franzini and W. Narendranathan (1982). "Serial Correlation and the Fixed Effects Model," *Review of Economic Studies*, 49:4, pp. 533-549.

Bleakley, Hoyt and Jeffrey C. Fuhrer (1997). "Job Matching, Shifts in the Beveridge Curve, and Labor Market Dynamics," *New England Economic Review*, September/October, pp. 3-19.

Bartik, Timothy J. (1991). *Who Benefits from Local Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004). "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, pp. 249-275 (February).

Card, David and Alan B. Krueger (1994). "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review*, 84:4, pp. 772-93.

Cameron, A. Colin, Jonah B. Gelbach and Douglas L. Miller (2006). "Robust Inference with Multi-Way Clustering," NBER Technical Working Paper No. 327 (September).

DellaVigna, Stefano and Joshua M. Pollet (2007, forthcoming). "Demograhics and Industry Returns," *American Economic Review.*

Driscoll, John and Aart Kraay (1998). "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data," *Review of Economics and Statistics*, pp. 549-560.

Elhorst, J. Paul (2003a). "Specification and Estimation of Spatial Panel Data Models," *International Regional Science Review*, 26:3, pp. 244-268.

___________ (2003b). "Dynamic Models in Space and Time," University of Groningen manuscript.

___________. (2003c). "Unconditional Maximum Likelihood Estimation of Dynamic Models for Spatial Panels," University of Groningen manuscript.

Foote, Christopher L. (2002). "Young Persons and State-Level Unemployment: The Role of the Housing Market," Harvard University manuscript.

Hansen, Christian (forthcoming *a*). "Generalized Least Squares Inference in Panel and Multilevel Model with Serial Correlation and Fixed Effects," *Journal of Econometrics*.

Hansen, Christian (forthcoming *b*). "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when T is Large," *Journal of Econometrics*.

Hoechle, Daniel (2007). "Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence," *Stata Journal*, 7:3, pp. 281-312.

Ingram, D.D.; Weed, J.A.; Parker, J.D.; Hamilton, B.; Schenker, N.; Arias, E.; and Madans J.H. (2003). "United States Census 2000 Population with Bridged Race Categories," *Vital Health Statistics*, 2:135. Hyattsville, Maryland: National Center for Health Statistics.

Jaimovich, Nir and Henry E. Siu (2007). "The Young, the Old, and the Restless: Demographics and Business Cycle Volatility," Federal Reserve Bank of Minneapolis Research Department Staff Report No. 387.

Kézdi, Gábor (2004). "Robust Standard Error Estimation in Fixed-Effects Panel Models," *Hungarian Statistical Review,* Special English Volume No. 9, pp. 95-116.

Lee, Lung-fei and Jihai Yu (2007). "A Spatial Dynamic Panel Data Model with Both Time and Individual Fixed Effects," The Ohio State University manuscript.

Nickell, Stephen (1981). "Biases in Dynamic Models with Fixed Effects," *Econometrica* 49:6, pp. 1417-1426.

Shimer, Robert (1998). "Why is the U.S. Unemployment Rate So Much Lower?" *NBER Macroeconomics Annual,* pp. 11-74.

Shimer, Robert (2001). "The Impact of Young Workers on the Aggregate Labor Market," *Quarterly Journal of Economics*, August, pp. 969-1008.

Solon, Gary (1984). "Estimating Autocorrelations in Fixed Effects Models," NBER Technical Working Paper No. 32.

Su, Liangjun and Zhenlin Yang (2007). "QML Estimation of Dynamic Panel Data Models with Spatial Errors," Singapore Management University manuscript.

Surveillance, Epidemiology, and End Results (SEER) Program Populations for 1969-2002 (2005). National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2005. (`www.seer.cancer.gov/popdata`)

Thompson, Samuel B. (2006). "Simple Formulas for Standard Errors that Cluster by Both Firm and Time," Harvard University mimeo.

Yu, Jihai, Robert de Jong, and Lung-fei Lee (2006). "Quasi-Maximum Likelihood Estimators for Spatial Dynamic Panel Data With Fixed Effects When Both n and T are Large," The Ohio State University manuscript.

Table I: Estimates of the Youth-Share Effect on State-Level Unemployment

	(1)	(2)
Sample Period	1973–1996	1973–2005
Panel A: OLS	-1.55	-.42
Robust Standard Error	(.18)**	(.14)**
Clustered by state	(.39)**	(.26)
Clustered by year	(.50)**	(.41)
Clustered by State and year	(.61)**	(.46)
Panel B: IV	-1.90	-.19
Robust Standard Error	(.25)**	(.17)
Clustered by state	(.53)**	(.31)
Clustered by Year	(.68)**	(.48)
Clustered by state and year	(.82)**	(.54)
Panel C: OLS-AR1	-1.36	-1.02
Robust Standard Error	(.31)**	(.26)**
Clustered by state	(.36)**	(.29)**
Clustered by year	(.73)*	(.55)*
Clustered by state and year	(.75)*	(.56)*
Panel D: IV-AR1	-1.68	-.82
Robust Standard Error	(.44)**	(.32)**
Clustered by state	(.45)**	(.41)**
Clustered by year	(1.02)	(.78)
Clustered by state and year	(1.02)	(.81)

Notes to Table I: The table presents various estimates of the youth-share coefficient from Equation (1) of the text, along with alternative standard errors. The dependent variable for each regression is the natural log of the unemployment rate for state i in year t. The coefficients in the table correspond to the natural log of the share of persons aged 16-64 who are aged 16-24, according to the SEER data (2005), which are adjustments of Census counts. All regressions include state and year dummies. The number of states in all regressions is 48 (AK, HI, and DC are always omitted). The data are balanced, so Column 1 has 24 years $\times$ 48 states = 1152 observations while Column 2 has 33 $\times$ 48 = 1584 observations. The instrument used for the youth share in Panels C and D is the log of the sum of lagged birth rates 16 to 24 years ago, as described in Shimer (2001). The AR1 parameters used to quasi-difference the data in Panels C and D are corrected by the method of Hansen (forthcoming a). In each regression in Panel D, the birth-rate instrument is also quasi-differenced. An asterisk (*) denotes significance at the 10% level; two asterisks (**) denote significance at 5%.

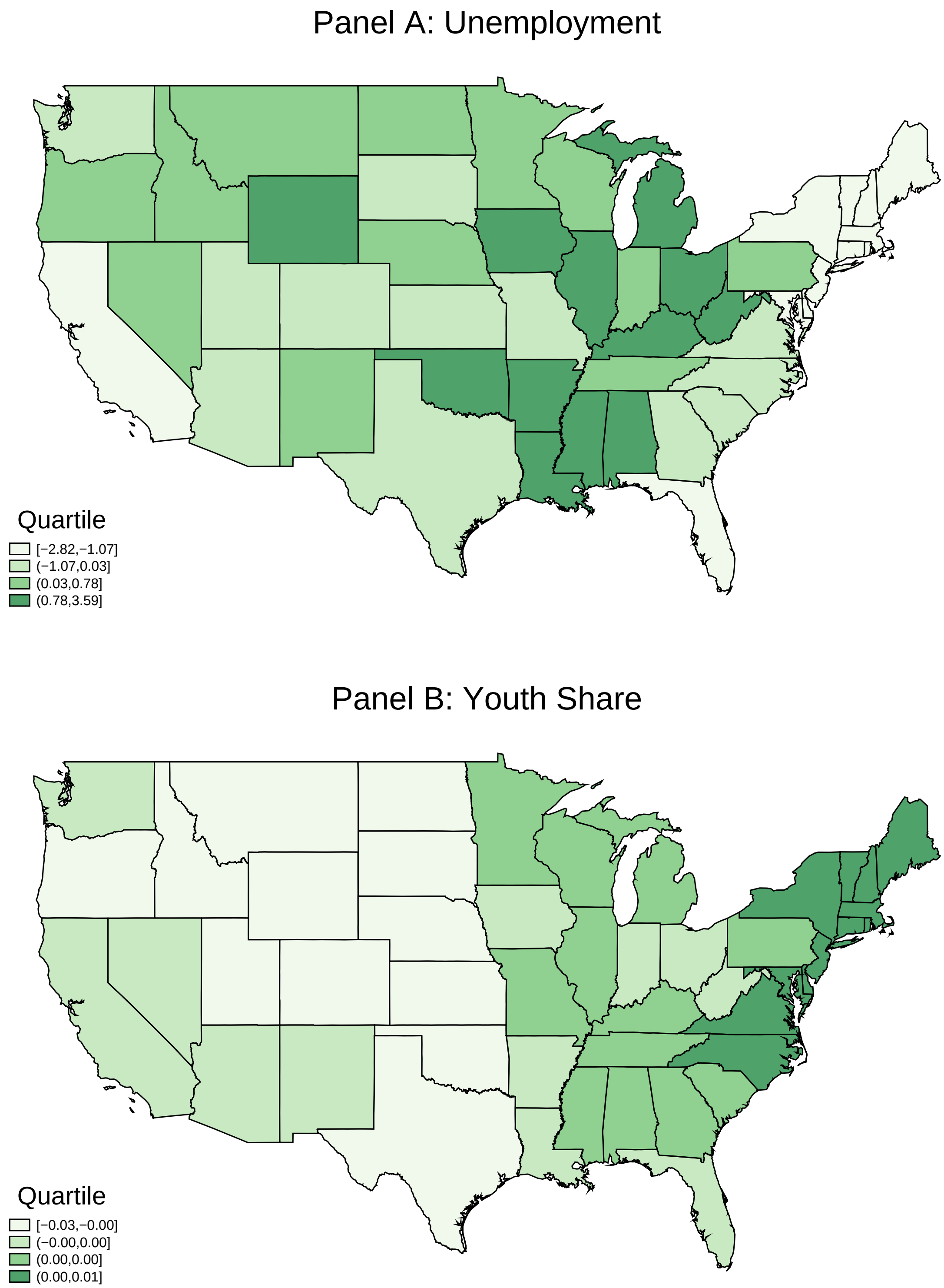

Figure 1: Spatial Correlation in Unemployment and Youth Shares in 1985. The data in each panel correspond to 1985 deviations from state and year means of the log of the unemployment rate (top panel) or the log of the youth share (bottom panel). The means are taken over the years 1973 to 2005 (the sample for Column 2 in Tables I and II).

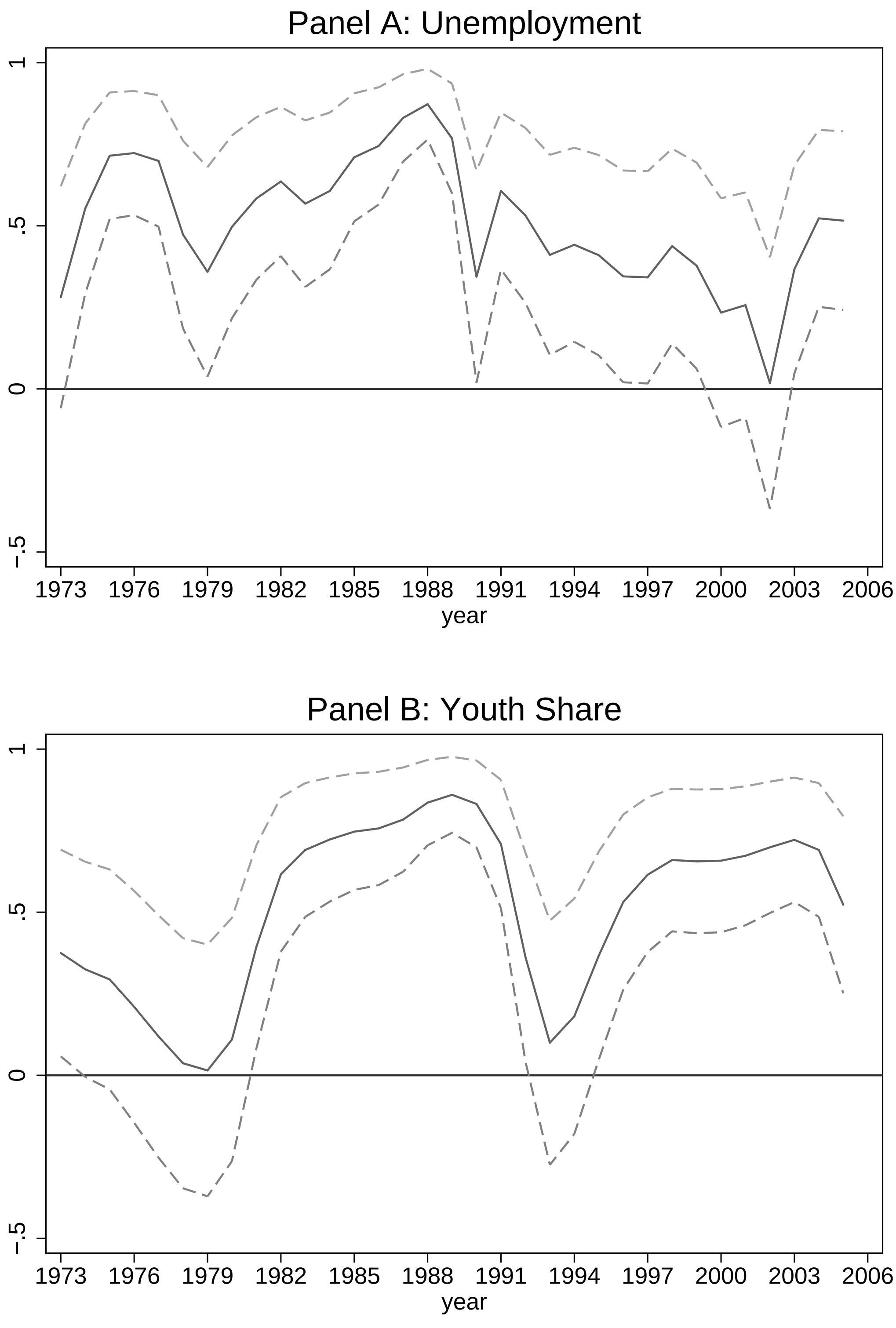

Figure 2: Spatial Correlation in Unemployment and Youth Shares: 1973–2005. Each panel presents estimates of the spatial correlation parameter (λ) from 33 separate spatial autocorrelation regressions. For a variable X from state i in a single year, this model is $X_i = \lambda W_N \mathbf{X} + v_i$, where $\mathbf{X}$ is the $N \times 1$ vector of X_i for all states, and the $N \times N$ weighting matrix W_N is a first-order contiguity matrix, with the $(i, j)th$ element equal to 1 if states i and j share a common border. The weighting matrix is row-normalized so that each row sums to 1. The data are deviated from state and year means before the separate spatial autocorrelation models are run. Dotted lines correspond to 95% confidence intervals.

Table II: Estimates of the Youth-Share Effect on State-Level Unemployment Allowing for Error Correlations Spanning Different States and Years

	(1)	(2)
Sample Period	1973–1996	1973–2005
Panel A: OLS	-1.55	-.42
Clustered by state & year	(.61)**	(.46)
Driscoll-Kraay, 1 lag	(.66)**	(.56)
Driscoll-Kraay, 3 lags	(.80)*	(.71)
DellaVigna & Pollet	(1.45)	(1.24)
Thompson, 1 lag	(.81)*	(.67)
Thompson, 3 lags	(.94)	(.82)
Panel B: IV	-1.90	-.19
Clustered by state & year	(.82)**	(.54)
Driscoll-Kraay, 1 lag	(.89)**	(.65)
Driscoll-Kraay, 3 lags	(1.09)*	(.83)
DellaVigna & Pollet	(1.99)	(1.46)
Thompson, 1 lag	(1.09)*	(.78)
Thompson, 3 lags	(1.26)	(.97)
Panel C: OLS-AR1	-1.36	-1.02
Clustered by state & year	(.75)*	(.56)*
Driscoll-Kraay, 1 lag	(.83)	(.62)
Driscoll-Kraay, 3 lags	(.89)	(.70)
DellaVigna & Pollet	(.80)*	(.60)
Thompson, 1 lag	(.89)	(.67)
Thompson, 3 lags	(.94)	(.78)
Panel D: IV-AR1	-1.68	-.82
Clustered by state & year	(1.02)	(.81)
Driscoll-Kraay, 1 lag	(1.25)	(.93)
Driscoll-Kraay, 3 lags	(1.42)	(1.08)
DellaVigna & Pollet	(1.12)	(.88)
Thompson, 1 lag	(1.40)	(1.03)
Thompson, 3 lags	(1.52)	(1.21)

Notes to Table II: The table presents various estimates of the youth-share coefficient in Equation (1) from the text (as does Table I), along with alternative standard errors. The coefficient estimates and the state-year clustered errors in the first row of each panel are taken directly from Table I; see the Notes to that table for a description of the variables used and the regression specification. An asterisk (*) denotes significance at the 10% level; two asterisks (**) denote significance at 5%.

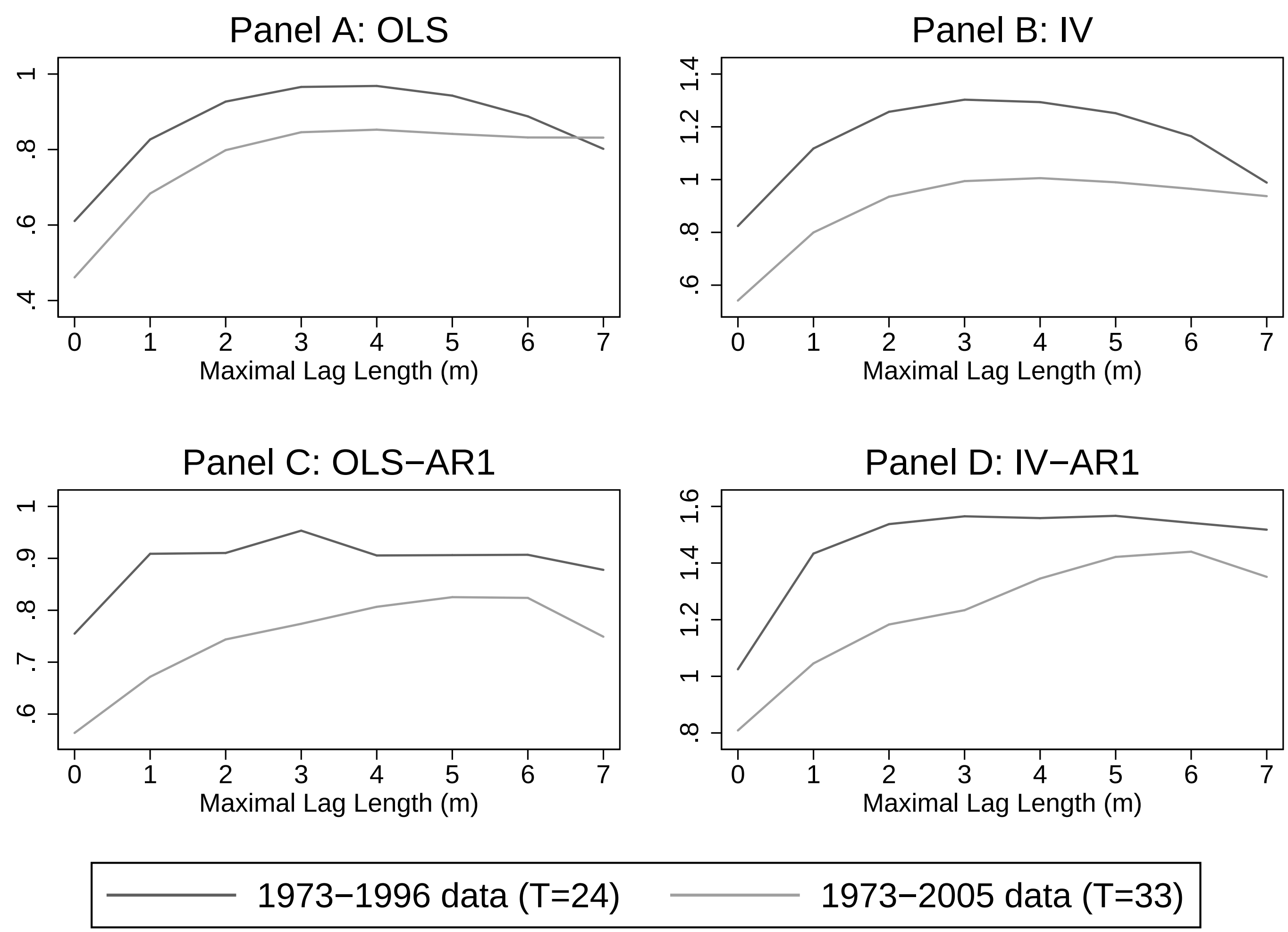

Figure 3: **Effect of Raising Maximal Lag Length on Thompson Standard Errors.** The lines in each panel correspond to the standard errors for the youth share coefficient in the state-level unemployment regression from Tables I and II. The darker lines correspond to the 1973–1996 sample (used for Column 1 of Tables I and II) while the lighter lines correspond to the 1973–2005 sample (used for Column 2). The value of m corresponds to the maximal lag length for which cross-state correlations are allowed to exist. A value $m = 0$ corresponds to standard errors clustered by state and year alone.

Table III: Estimates of the Youth-Share Effect on State-Level Unemployment: 1973–2005

	(1)	(2)	(3)	(4)	(5)	(6)
Bartik Variables Included as Regressors?	No	Yes	Yes	Yes	Yes	Yes
Geographic Interactions	None	None	Reg-Yr	Div-Yr	Div-Yr	Div-Yr
AR Correction	None	None	None	None	AR(1)	AR(2)
Panel A: OLS	-.42	-.29	-.04	-.17	-.58	-.51
Robust	(.14)**	(.12)**	(.13)	(.14)	(.27)**	(.27)*
Clustered by state	(.26)	(.23)	(.26)	(.25)	(.26)**	(.25)**
Clustered by year	(.41)	(.33)	(.24)	(.20)	(.35)	(.35)
Clustered by state & year	(.46)	(.38)	(.33)	(.29)	(.35)*	(.34)
Thompson, 1 lag	(.67)	(.55)	(.41)	(.34)	(.39)	(.38)
Thompson, 3 lags	(.82)	(.65)	(.49)	(.38)	(.40)	(.39)
Panel B: IV	-.19	-.21	.30	.35	.21	.30
Robust	(.17)	(.15)	(.17)*	(.21)*	(.45)	(.44)
Clustered by state	(.31)	(.29)	(.40)	(.44)	(.53)	(.50)
Clustered by year	(.48)	(.39)	(.33)	(.34)	(.61)	(.61)
Clustered by state & year	(.54)	(.46)	(.49)	(.51)	(.67)	(.66)
Thompson, 1 lag	(.78)	(.65)	(.60)	(.58)	(.70)	(.69)
Thompson, 3 lags	(.97)	(.78)	(.68)	(.64)	(.76)	(.77)

Notes to Table III: The table presents various estimates of the youth-share coefficient from equation (1') of the text, estimated on the 1973–2005 sample. See the notes to that table for descriptions of the variables used and the regression specification. Panel B uses the birth-rate instrument from Tables I and II as an IV for the youth-share. The Bartik variable used in Columns 2–6 is constructed by weighting national industry-level growth rates by state-specific industry weights from various base years. The "Reg-Yr" interactions correspond to interactions between dummy variables for (three of the) four Census regions of the country; the "Div-Yr" interactions are similarly constructed using the nine Census divisions. The AR parameters used to quasi-difference the data in Columns 5 and 6 are corrected by the method of Hansen (forthcoming *a*). The birth-rate instrument is also quasi-differenced in the Columns 5 and 6 of Panel B. An asterisk (*) denotes significance at the 10% level; two asterisks (**) denote significance at 5%.